



OIPE

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/023,929DATE: 02/10/2002
TIME: 14:53:36Input Set : N:\Cr3\RULE60\10023929.txt
Output Set: N:\CRF3\02102002\J023929.rawRECEIVED
FEB 21 2002
TECH CENTER 1600/2900

SEQUENCE LISTING

C--> 5 (1) GENERAL INFORMATION:
7 (i) APPLICANT: Hillman, Jennifer L.
8 Goli, Surya K.
C--> 10 (ii) TITLE OF INVENTION: NOVEL HUMAN MLS3 PROTEIN
13 (iii) NUMBER OF SEQUENCES: 6
15 (iv) CORRESPONDENCE ADDRESS:
16 (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
17 (B) STREET: 3174 Porter Drive
18 (C) CITY: Palo Alto
19 (D) STATE: CA
20 (E) COUNTRY: USA
21 (F) ZIP: 94304
23 (v) COMPUTER READABLE FORM:
24 (A) MEDIUM TYPE: Diskette
25 (B) COMPUTER: IBM Compatible
26 (C) OPERATING SYSTEM: DOS
27 (D) SOFTWARE: FastSEQ for Windows Version 2.0
34 (vi) CURRENT APPLICATION DATA:
C--> 35 (A) APPLICATION NUMBER: US/10/023,929
C--> 36 (B) FILING DATE: 17-Dec-2001
37 (C) CLASSIFICATION:
40 (vii) PRIOR APPLICATION DATA:
41 (A) APPLICATION NUMBER: 08/805,965
42 (B) FILING DATE:
44 (viii) ATTORNEY/AGENT INFORMATION:
45 (A) NAME: Billings, Lucy J.
46 (B) REGISTRATION NUMBER: 36,749
47 (C) REFERENCE/DOCKET NUMBER: PF-0223 US
49 (ix) TELECOMMUNICATION INFORMATION:
50 (A) TELEPHONE: 415-855-0555
51 (B) TELEFAX: 415-845-4166
52 (C) TELEX:
55 (2) INFORMATION FOR SEQ ID NO: 1:
57 (i) SEQUENCE CHARACTERISTICS:
58 (A) LENGTH: 262 amino acids
59 (B) TYPE: amino acid
60 (C) STRANDEDNESS: single
61 (D) TOPOLOGY: linear
63 (vii) IMMEDIATE SOURCE:
64 (A) LIBRARY: BRAITUT02
65 (B) CLONE: 762280
67 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/023,929

DATE: 02/10/2002

TIME: 14:53:36

Input Set : N:\Crf3\RULE60\10023929.txt

Output Set: N:\CRF3\02102002\J023929.raw

```

69 Met Phe Arg Phe Met Arg Asp Val Glu Pro Glu Asp Pro Met Phe Leu
70 1 5 10 15
71 Met Asp Pro Phe Ala Ile His Arg Gln His Met Ser Arg Met Leu Ser
72 20 25 30
73 Gly Gly Phe Gly Tyr Ser Pro Phe Leu Ser Ile Thr Asp Gly Asn Met
74 35 40 45
75 Pro Gly Thr Arg Ala Ala Ser Arg Arg Met Gln Gln Ala Gly Ala Val
76 50 55 60
W--> 77 Xaa Pro Phe Gly Xaa Leu Gly Met Ser Gly Gly Phe Met Asp Met Phe
78 65 70 75 80
W--> 79 Gly Met Met Asn Asp Met Xaa Gly Asn Met Glu His Met Thr Ala Gly
80 85 90 95
81 Gly Asn Cys Gln Thr Phe Ser Ser Ser Thr Val Ile Ser Tyr Ser Asn
82 100 105 110
83 Thr Gly Asp Gly Ala Pro Lys Val Tyr Gln Glu Thr Ser Glu Met Arg
84 115 120 125
85 Ser Ala Pro Gly Gly Ile Arg Glu Thr Arg Arg Thr Val Arg Asp Ser
86 130 135 140
87 Asp Ser Gly Leu Glu Gln Met Ser Ile Gly His His Ile Arg Asp Arg
88 145 150 155 160
89 Ala His Ile Leu Gln Arg Ser Arg Asn His Arg Thr Gly Asp Gln Glu
90 165 170 175
91 Glu Arg Gln Asp Tyr Ile Asn Leu Asp Glu Ser Glu Ala Ala Ala Phe
92 180 185 190
93 Asp Asp Glu Trp Arg Arg Glu Thr Ser Arg Phe Arg Gln Gln Arg Pro
94 195 200 205
95 Leu Glu Phe Arg Arg Leu Glu Ser Ser Gly Ala Gly Gly Arg Arg Ala
96 210 215 220
97 Glu Gly Pro Pro Arg Leu Ala Ile Gln Gly Pro Glu Asp Ser Leu Pro
98 225 230 235 240
99 Asp Ser Pro Ala Ala Met Thr Gly Glu Gly Pro Gly Ala Ser Ala Leu
100 245 250 255
101 Leu Tyr Arg Leu Arg Gly
102 260
104 (2) INFORMATION FOR SEQ ID NO: 2:
106 (i) SEQUENCE CHARACTERISTICS:
107 (A) LENGTH: 1322 base pairs
108 (B) TYPE: nucleic acid
109 (C) STRANDEDNESS: single
110 (D) TOPOLOGY: linear
112 (vii) IMMEDIATE SOURCE:
113 (A) LIBRARY: BRAITUT02
114 (B) CLONE: 762280
116 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
118 GGGGGGCGTA CGGAGGTGGC AGCTGTGGGA GGAGGCGGCG TGGAAGGCCG AGGAGCTCAA 60
119 GCCCGGACCA ATCCCCACGT TCCGGGCCGC CACCCTGACC CTGCAGCGTA CCGGGAAGCG 120
120 AAACCGGCCG GATGGGCCGC TGAGCCCGAA TCGGGCACTG TGTGGAGCCC CCTGGAGCTG 180
121 AGATCAGGAT GTTCGCTTC ATGAGGGACG TGGAGCCTGA GGATCCCATG TTCCTGATGG 240
122 ATCCCTTTC TATTCACCGT CAGCATATGA GCCGTATGTT GTCAGGTGGC TTTGGATATA 300

```

RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/10/023,929

TIME: 14:53:36

Input Set : N:\Crif3\RULE60\10023929.txt

Output Set: N:\CRF3\02102002\J023929.raw

```

123  GCCCCCTTCCT CAGCATCACA GATGGCAACA TGCCAGGGAC CAGGGCTGCC AGCCGCCGGA 360
124  TGCAGCAGGC TGGAGCTGTC TNCCCCTTTG GGNTGCTGGG AATGTCGGGT GGTTTCATGG 420
125  ACATGTTTGG GATGATGAAT GACATGNTTG GAAACATGGA ACACATGACA GCTGGAGGCA 480
126  ATTGCCAGAC CTTCTCATCT TCCACTGTCA TCTCCTACTC CAATACGGGT GATGGTGCCC 540
127  CCAAGGTCTA CCAAGAGACA TCAGAGATGC GCTCGGCACC AGGCGGGATC CGGGAGACAC 600
128  GGAGGACTGT TCGGGATTCA GACAGTGGAC TGGAGCAGAT GTCCATTGGG CATCACATCC 660
129  GGGACAGGGC TCACATCCTC CAGCGCTCCC GAAACCATCG CACGGGGGAC CAGGAGGAGC 720
130  GGCAGGACTA TATCAACCTG GATGAGAGTG AGGCCGCAGC GTTTGATGAC GAGTGGCGGC 780
131  GGGAGACCTC CCGATTCCGG CAGCAGCGTC CCCTGGAGTT TCGGCGGCTT GAGTCTCTAG 840
132  GGGCTGGGGG ACGAAGGGCG GAGGGGCCTC CCCGCCTGGC CATCCAGGGA CCTGAGGACT 900
133  CCCTTCCCGA CAGTCCCGCC GCTATGACTG GTGAGGGCCC CGGGGCCTCA GCTCTCTTGT 960
134  ACAGGCTGAG AGGCTGAGAA ATCATCCCCT GAATAACTTT TTCCTCTCGA TTCCCATCCC 1020
135  CAATTTAATA TTAAATTAAC AGGCAAGCCG GCCCCACCT CTCCCTGGGG GTCTCAGGGA 1080
136  GAACCTTTCA CGGCACCCTT TCCCTACCTT TTCCTTCTTT AATCTCCTGG TTTACCATTG 1140
137  ATGACTTCGG CTCTGCATCT ACTTACTTGA TTTTTCATTC TGCCACTTCA TCTTCAAACC 1200
138  CCCTCACCTT TCCCATCCTA CTCCTGCCAT GCATTGAAGG GTCAATGCAT TTTGGGGTGA 1260
139  GNTTNGGTTT AGGGGCCCCCT TCATNCCTNA GCTACCTGGG TCTTTGCCCA ACTTTTCTCA 1320
140  GA 1322

```

142 (2) INFORMATION FOR SEQ ID NO: 3:

144 (i) SEQUENCE CHARACTERISTICS:

145 (A) LENGTH: 268 amino acids

146 (B) TYPE: amino acid

147 (C) STRANDEDNESS: single

148 (D) TOPOLOGY: linear

150 (vii) IMMEDIATE SOURCE:

151 (A) LIBRARY: GenBank

152 (B) CLONE: 1066392

154 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

156 Met Phe Arg Met Leu Asn Ser Ser Phe Glu Asp Asp Pro Phe Phe Ser
157 1 5 10 15
158 Glu Ser Ile Leu Ala His Arg Glu Asn Met Arg Gln Met Ile Arg Ser
159 20 25 30
160 Phe Ser Glu Pro Phe Gly Arg Asp Leu Leu Ser Ile Ser Asp Gly Arg
161 35 40 45
162 Gly Arg Ala His Asn Arg Arg Gly His Asn Asp Gly Glu Asp Ser Leu
163 50 55 60
164 Thr His Thr Asp Val Ser Ser Phe Gln Thr Met Asp Gln Met Val Ser
165 65 70 75 80
166 Asn Met Arg Asn Tyr Met Gln Lys Leu Glu Arg Asn Phe Gly Gln Leu
167 85 90 95
168 Ser Val Asp Pro Asn Gly His Ser Phe Cys Ser Ser Ser Val Met Thr
169 100 105 110
170 Tyr Ser Lys Ile Gly Asp Glu Pro Pro Lys Val Phe Gln Ala Ser Thr
171 115 120 125
172 Gln Thr Arg Arg Ala Pro Gly Gly Ile Lys Glu Thr Arg Lys Ala Met
173 130 135 140
174 Arg Asp Ser Asp Ser Gly Leu Glu Lys Met Ala Ile Gly His His Ile
175 145 150 155 160
176 His Asp Arg Ala His Val Ile Lys Lys Ser Lys Asn Lys Lys Thr Gly

```

RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/10/023,929

TIME: 14:53:36

Input Set : N:\Crf3\RULE60\10023929.txt

Output Set: N:\CRF3\02102002\J023929.raw

```

177          165          170          175
178 Asp Glu Glu Val Asn Gln Glu Phe Ile Asn Met Asn Glu Ser Asp Ala
179          180          185          190
180 His Ala Phe Asp Glu Glu Trp Gln Ser Glu Val Leu Lys Tyr Lys Pro
181          195          200          205
182 Gly Arg His Asn Leu Gly Asn Thr Arg Met Arg Ser Val Gly His Glu
183          210          215          220
184 Asn Pro Gly Ser Arg Glu Leu Lys Arg Arg Glu Lys Pro Gln Gln Ser
185          225          230          235          240
186 Pro Ala Ile Glu His Gly Arg Arg Ser Asn Val Leu Gly Asp Lys Leu
187          245          250          255
188 His Ile Lys Gly Ser Ser Val Lys Ser Asn Lys Lys
189          260          265

```

191 (2) INFORMATION FOR SEQ ID NO: 4:

193 (i) SEQUENCE CHARACTERISTICS:

194 (A) LENGTH: 1116 base pairs

195 (B) TYPE: nucleic acid

196 (C) STRANDEDNESS: single

197 (D) TOPOLOGY: linear

199 (vii) IMMEDIATE SOURCE:

200 (A) LIBRARY: GenBank

201 (B) CLONE: 1066391

203 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

205 GTTATGTGTT CCCGTCCGTA CTGGAGGCTA GCTCTTGTCG CGGCCCGCGGC GAGTTAACAT      60
206 CGTTTTTCCA ATCTGTCCGC GGCTGCCGCC ACCCAAGACA GAGCCAGAAT GTTCAGGATG      120
207 CTGAACAGCA GTTTTGAGGA TGACCCCTTC TTCTCTGAGT CCATTCTTGC ACACCGAGAA      180
208 AATATGCGAC AGATGATAAG AAGTTTTTCT GAACCCCTTTG GAAGAGACTT GCTCAGTATC      240
209 TCTGATGGTA GAGGGAGAGC TCATAATCGT AGAGGACATA ATGATGGTGA AGATTCTTTG      300
210 ACTCATACAG ATGTCAGCTC TTTCCAGACC ATGGACCAAA TGGTGTCAAA TATGAGAAAC      360
211 TATATGCAGA AATTAGAAAG AAACCTCGGT CAACTTTCAG TGGATCCAAA TGGACATTCA      420
212 TTTTGTCTCT CCTCAGTTAT GACTTATTCC AAAATAGGAG ATGAACCGCC AAAGGTTTTT      480
213 CAGGCCTCAA CTCAAACCTG TCGAGCTCCA GGAGGAATAA AGGAAACCAG GAAAGCAATG      540
214 AGAGATTCTG ACAGTGGACT AGAAAAAATG GCTATTGGTC ATCATATCCA TGACCGAGCT      600
215 CATGTCATTA AAAAGTCAAA GAACAAGAAG ACTGGAGATG AAGAGGTCAA CCAGGAGTTC      660
216 ATCAATATGA ATGAAAAGCGA TGCTCATGCT TTTGATGAGG AGTGGCAAAG TGAGGTTTTG      720
217 AAGTACAAAC CAGGACGACA CAATCTAGGA AACACTAGAA TGAGAAGTGT TGGCCATGAG      780
218 AATCCTGGCT CCCGAGAACT TAAAAGAAGG GAGAAACCTC AACAAAGTCC AGCCATTGAA      840
219 CATGGAAGGA GATCAAATGT TTTGGGGGAC AAACCTCCACA TCAAAGGCTC ATCTGTGAAA      900
220 AGCAACAAAA AATAAATAGC CATGCATTG ATTTGTTTAG TTTTGATTGT TTTAACAGTT      960
221 AGTAATGGTG CTGGGTAATA AGCATAAGAC CAATCTCTTG CTGTTAAATC AGTTCTGTCC      1020
222 TTGGCAACTT TCTTCTGATA TCTGAATGTT CATGAAGGTC CTAGCTTTAT ATTGTCCCTC      1080
223 TTTTAGGAAT AAAATTTTGA TTTTCAACAA AAAAAA      1116

```

225 (2) INFORMATION FOR SEQ ID NO: 5:

227 (i) SEQUENCE CHARACTERISTICS:

228 (A) LENGTH: 248 amino acids

229 (B) TYPE: amino acid

230 (C) STRANDEDNESS: single

231 (D) TOPOLOGY: linear

233 (vii) IMMEDIATE SOURCE:

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/023,929

DATE: 02/10/2002

TIME: 14:53:36

Input Set : N:\Crf3\RULE60\10023929.txt

Output Set: N:\CRF3\02102002\J023929.raw

```

234      (A) LIBRARY: GenBank
235      (B) CLONE: 1399745
237      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
239  Met Phe Arg Phe Met Arg Asp Val Glu Pro Glu Asp Pro Met Phe Leu
240    1          5          10          15
241  Met Asp Pro Phe Ala Ile His Arg Gln His Met Ser Arg Met Leu Ser
242          20          25          30
243  Gly Gly Phe Gly Tyr Ser Pro Phe Leu Ser Ile Thr Asp Gly Asn Met
244          35          40          45
245  Pro Gly Thr Arg Pro Ala Ser Arg Arg Met Gln Gln Ala Gly Ala Val
246          50          55          60
247  Ser Pro Phe Gly Met Leu Gly Met Ser Gly Gly Phe Met Asp Met Phe
248          65          70          75          80
249  Gly Met Met Asn Asp Met Ile Gly Asn Met Glu His Met Thr Ala Gly
250          85          90          95
251  Gly Asn Cys Gln Thr Phe Ser Ser Ser Thr Val Ile Ser Tyr Ser Asn
252          100         105         110
253  Thr Gly Asp Gly Ala Pro Lys Val Tyr Gln Glu Thr Ser Glu Met Arg
254          115         120         125
255  Ser Ala Pro Gly Gly Ile Arg Glu Thr Arg Arg Thr Val Arg Asp Ser
256          130         135         140
257  Asp Ser Gly Leu Glu Gln Met Ser Ile Gly His His Ile Arg Asp Arg
258          145         150         155         160
259  Ala His Ile Leu Gln Arg Ser Arg Asn His Arg Thr Gly Asp Gln Glu
260          165         170         175
261  Glu Arg Gln Asp Tyr Ile Asn Leu Asp Glu Ser Glu Ala Ala Ala Phe
262          180         185         190
263  Asp Asp Glu Trp Arg Arg Glu Thr Ser Arg Phe Arg Gln Gln Arg Pro
264          195         200         205
265  Leu Glu Phe Arg Arg Leu Glu Ser Ser Gly Ala Gly Gly Arg Arg Ala
266          210         215         220
267  Glu Gly Pro Pro Arg Leu Ala Ile Gln Gly Pro Glu Asp Ser Pro Ser
268          225         230         235         240
269  Arg Gln Ser Arg Arg Tyr Asp Trp
270          245
272 (2) INFORMATION FOR SEQ ID NO: 6:
274      (i) SEQUENCE CHARACTERISTICS:
275          (A) LENGTH: 1502 base pairs
276          (B) TYPE: nucleic acid
277          (C) STRANDEDNESS: single
278          (D) TOPOLOGY: linear
280      (vii) IMMEDIATE SOURCE:
281          (A) LIBRARY: GenBank
282          (B) CLONE: 1399744
284      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
286  CTCTAAAGGG CAGCTGTGGG AGGAGGCGGC GTGGAAGGCC GAGGAGCTCA AGCCCGGACC      60
287  AATCCCCACG TTCCGGGCCG CGACCCTGAC CCTGCAGCGT ACCGGGAAGC GAAACCGGCC      120
288  GGATGGGCCG CTGAGCCCGA ATCGGGCACT GTGTGGAGCC CCCTGGAGCT GAGATCAGGA      180
289  TGTTCGCTT CATGAGGGAC GTGGAGCCTG AGGATCCCAT GTTCCTGATG GATCCCTTTG      240

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/023,929

DATE: 02/10/2002

TIME: 14:53:37

Input Set : N:\Crf3\RULE60\10023929.txt

Output Set: N:\CRF3\02102002\J023929.raw

L:5 M:220 C: Keyword misspelled or invalid format, [(1) GENERAL INFORMATION:]
L:10 M:220 C: Keyword misspelled or invalid format, [(ii) TITLE OF INVENTION:]
L:30 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:31 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:35 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:36 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:77 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:79 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1